

**The Internal Revenue Service Could Improve
Its Process to More Reliably Measure the
Accuracy of Its Toll-Free Tax Law Assistance**

February 2002

Reference Number: 2002-40-051

This report has cleared the Treasury Inspector General for Tax Administration disclosure review process and information determined to be restricted from public release has been redacted from this document.



DEPARTMENT OF THE TREASURY
WASHINGTON, D.C. 20220

INSPECTOR GENERAL
for TAX
ADMINISTRATION

February 15, 2002

MEMORANDUM FOR COMMISSIONER, WAGE AND INVESTMENT DIVISION

Pamela J. Gardiner

FROM: Pamela J. Gardiner
Deputy Inspector General for Audit

SUBJECT: Final Audit Report – The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance (Audit # 200140039)

This report presents the results of our review of the Internal Revenue Service's (IRS) efforts to measure the accuracy of its toll-free tax law assistance. The overall objective of this review was to determine if the IRS reliably measured the accuracy of responses that millions of taxpayers experienced when they called the IRS for tax law assistance.

In summary, we found that the IRS made a good attempt to obtain a representative measure of its tax law assistance within the limitations of its current sampling strategy. The IRS implemented a sampling strategy that required quality reviewers to select, monitor, and assess the quality of responses to calls in "real" time during live calls. During the period that it monitored the calls, the results provided useful indicators of the quality of the responses it provided. However, weaknesses in the current sampling strategy had the potential to bias the accuracy of the results. For example:

- The IRS' sampling strategy used a cluster sampling design, an acceptable statistically valid random sampling method. However, the sampling plan developed from the design may not have produced a valid random sample. The IRS' sampling plan attempted to ensure a certain number of calls were reviewed from all call sites¹ at various times of the day; however, the plan did not include all hours and days of operation, used a sampling formula that likely resulted in a

¹ A call site is any location within the IRS where quantities of incoming and outgoing calls are handled by people, telephones, and computers. The IRS' call sites provide assistance to taxpayers by answering questions, providing assistance, and resolving account-related issues.

wider precision than reported, was designed in a mutually exclusive manner (i.e., if one event occurs, the other(s) cannot), and did not include r-mail² calls.

- The sampling plan created by the IRS was not implemented as designed. To ensure the reliability of the results, once a random sampling plan has been designed, it is critical that it be followed to achieve the expected outcome. Our review found several problems in this area, including latitude on the part of the reviewers to select calls to monitor. By deviating from the sampling plan and disconnecting from calls, reviewers had the ability to influence which calls were included in the sample, which may distort the true randomness of the sample. In addition, the required number of calls was not reviewed at critical times during the 2001 Filing Season.

We were unable to assess the impact of these weaknesses on the 73.78 percent accuracy rate the IRS reported for the 2001 Filing Season. (The IRS' true accuracy for toll-free tax law assistance could be higher or lower than estimated or at 73.78 percent as reported.) To do this would require the re-creation of the sample, which is not possible.

Also, managerial reviews were not conducted as required at critical periods during the 2001 Filing Season. In addition, they were not conducted in an independent setting. Improving this process could reduce some of the weaknesses identified in the current sampling strategy, such as deviating from the sampling plan and disconnecting from calls.

Management's Response: IRS management agreed with the findings we presented in our report and implemented corrective actions where possible. They have already implemented corrective actions for recommendations 1 and 2 of our report. However, they were not able to devise corrective actions for recommendations 3 and 4.

Specifically, IRS management developed a sampling plan for the 2002 Filing Season that covers all of the days and hours Customer Service Representatives will provide tax law service. They also hired additional Centralized Quality Review Site (CQRS) staff and authorized overtime to assure the IRS executes the sample plan as designed. The new sampling plan offers the highest level of coverage by using the most random sampling its telephone system allows.

In addition, the 2002 version of the plan does not allow reviewers to select call sites and applications. Instead, reviewers are assigned to specific sites and applications. IRS management plans to monitor the samples to ensure reviewers are not deviating from their hourly assignments and that projected sample sizes are attained. The IRS has scheduled staff resources, including overtime, to ensure it consistently meets sample sizes.

² R-mail includes questions answered by return telephone call or electronic mail.

However, the IRS was not able to develop corrective actions for recommendations 3 or 4 and one aspect of recommendation 2. Specifically, the IRS could not implement recommendation number 3 due to telephone system limitations. The telephone system used by CQRS does not allow simultaneous remote call monitoring. The IRS did strengthen controls to ensure that each CQRS manager meets minimum monitoring standards.

In addition, while IRS management agrees that an automated call recording system is the best way to achieve random call selection, they pointed out that the IRS is not technologically able to do so now. The IRS is aggressively pursuing automated call recording. Lastly, the telephone equipment the CQRS staff uses cannot track disconnected calls.

Management's complete response to the draft report is included as Appendix VI.

Copies of this report are also being sent to the IRS managers who are affected by the report recommendations. Please contact me at (202) 622-6510 if you have questions or Michael R. Phillips, Acting Assistant Inspector General for Audit (Wage and Investment Income Programs), at (202) 927-7085.

**The Internal Revenue Service Could Improve Its Process to More
Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance**

Table of Contents

Background	Page 1
Weaknesses in the Sampling Strategy May Bias the Accuracy of the Results	Page 5
<u>Recommendations 1 and 2:</u>	Page 14
<u>Recommendations 3 and 4:</u>	Page 15
Appendix I – Detailed Objective, Scope, and Methodology	Page 17
Appendix II – Major Contributors to This Report	Page 19
Appendix III – Report Distribution List	Page 20
Appendix IV – Tables and Charts	Page 21
Appendix V – Common Sampling Techniques	Page 27
Appendix VI – Management’s Response to the Draft Report	Page 28

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

Background

Millions of taxpayers rely on the Internal Revenue Service (IRS) to provide accurate information when they call with questions about the tax law. Incorrect information, if provided by the IRS, could result in filing errors and increased taxpayer burden. During the 2001 Filing Season,¹ the IRS received approximately 4.7 million² toll-free telephone calls from taxpayers seeking tax law assistance. The assistance was provided by Customer Service Representatives (CSR) located in IRS call sites³ across the country. To assess the quality of this assistance, the IRS reviewed a sample of live taxpayer calls. For the 2001 Filing Season, it reported a national accuracy rate of 73.78⁴ percent for its toll-free tax law assistance that covered the 4.7 million calls.

Taxpayers calling the IRS for tax law assistance initially reached an automated, menu-driven telephone system and chose to either listen to recorded information or speak to a CSR. When taxpayers chose to speak to a CSR, they navigated through the system and selected a particular tax topic (application) that related to his or her question. For example, one application covered questions on filing status and dependents and another covered questions on pensions and social security benefits. The telephone system was designed to then route the call to the next available CSR trained in the application selected.

Incoming calls were routed to a CSR in any 1 of 15 call sites nationwide that handled tax law calls. Appendix IV

¹ The 2001 Filing Season refers to the period from January 2 to April 16, 2001. (January 1 was a holiday.)

² The 4.7 million calls relating to tax law did not include approximately 1 million e-mail calls and approximately 584,000 Spanish assistance calls. E-mail includes questions answered by return telephone call or electronic mail.

³ A call site is any location within the IRS where quantities of incoming and outgoing calls are handled by people, telephones, and computers. The IRS' call sites provide assistance to taxpayers by answering questions, providing assistance, and resolving account-related issues.

⁴ This percentage was at a 90 percent confidence level with a precision margin of 0.73 percent. This means that the IRS was 90 percent confident that nationwide it correctly answered toll-free tax law questions between 73.05 and 74.51 percent of the time during the filing season.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

(Figures 1 and 2) provides the locations of the tax law call sites and their hours of operation during the 2001 Filing Season. The volume of calls handled by the IRS varied by application and call site. Appendix IV (Figures 3 and 4) provides the volumes of calls handled by application and call site during the 2001 Filing Season.

An overview of the IRS' centralized quality review process for tax law assistance

The IRS established the Centralized Quality Review Site (CQRS)⁵ to determine the accuracy of responses provided to taxpayers that call the IRS' toll-free customer service telephone system with questions about the tax law. It would be impossible for the IRS to review every call for accuracy; therefore, the CQRS reviewers monitored a sample of taxpayer calls and determined the accuracy of the responses.

A sampling plan was developed by the IRS to select actual taxpayer calls to monitor. Using this sampling plan, reviewers selected and monitored the calls to determine whether the CSRs followed all procedures and provided the taxpayer with an accurate response to his or her question. The reviewers then annotated the results of their monitoring on a case review form. Information on the case review forms was entered into a computerized database from which reports were generated providing accuracy rates and other quality-related information.

To ensure that reviewers possessed the required skills to adequately monitor and assess the accuracy of the call, the IRS selected individuals that already had years of tax law experience. Reviewers were not expected to be experts on all tax law topics. They were assigned to monitor only two to four tax law applications. The IRS ensured the reviewers maintained the necessary skills by providing refresher training.

Sampling techniques

As stated earlier, it was not feasible for the IRS to monitor all calls received to gauge the accuracy of responses.

⁵ The site is located in Philadelphia, Pennsylvania.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

However, to analyze the quality of the responses, the IRS needed to collect data that were representative of all tax law calls received. Sampling offered the practical solution.

Sampling is the selection and study of a part of a whole (the population) for the purpose of drawing conclusions about the whole. Sampling may be likened to taste testing, where the tester tastes a small part of the item and thereby draws conclusions about the whole regarding its quality or other characteristics.⁶

In designing a sampling plan, there are a variety of possible sample selection techniques that can be used, each having various levels of reliability. Some commonly used techniques are random, judgmental, and convenience sampling. Appendix V provides descriptions of these sampling techniques.

The IRS chose the statistical method of random selection using cluster sampling. Cluster sampling randomly selects groups or clusters of units to sample from. One benefit of cluster sampling is the reduced cost of the sample selection and data collection.

One major benefit of statistical random sampling in general is that it permits objective measurement of the reliability of the results.⁷ Established mathematical formulas are used to determine the proper sample size necessary to measure the reliability of the results. This sampling method is used when conclusions about the entire population are to be presented.⁸

With all statistical random sampling methods, *every* item in the population is given an *equal* chance of being included in the sample.⁷ For example, each call answered by the IRS within the cluster selected must have an equal chance of being selected for the sample, or the method is not truly random. Unless each item is selected by a purely random technique, there is no way of measuring later how

⁶ Dr. B.J. Mandel & Dr. Robert E. Laessig, Statistics for Management, Dangary Publishing Company, Baltimore, MD., 1996, p. 174.

⁷ Mandel & Laessig, p. 204.

⁸ Mandel & Laessig, p. 203.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

accurately the sample reflects the characteristics of the population from which it was selected.

Sampling reliability

Since a sample only contains partial data, the results have some limitations because one can never be certain the entire population is represented. The limitations are identified through a process called “measurement of reliability.”

When designing a sampling plan, management must decide the desired reliability of the results and degree of confidence in using those results.⁹

Assessing the reliability of the results involves determining how closely the sample represents the population (precision margin) and the confidence with which the results can be used to assess the population (confidence level). Precision is the amount of error that management will tolerate due to sampling and is expressed as a plus or minus figure (such as +/-5 percent). The smaller the precision margin, the less error due to sampling exists.

The confidence level refers to the degree of assurance that the results obtained from the sample, after applying the precision margin, represent the true population. For illustration, the IRS designed its sample with a confidence level of 90 percent, a precision margin of +/-5 percent, and a historical accuracy rate of 71 percent. That means that the IRS would expect that the average accuracy rate for the calls sampled would fall within a range from 66 percent to 76 percent, 90 percent of the time.

How the IRS used the results from its sample

The IRS designed the sample to provide an estimate of the accuracy of its responses to tax law questions on both a nationwide and a call site level. In addition, the IRS used the results of the sample to identify needed improvements in training and research materials for CSRs and to propose changes to tax forms, instructions, and publications. The results of the sample could also enable managers to better

⁹ Mandel & Laessig, p. 28 and 181.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

operate their call sites and set performance goals to improve its telephone customer service program overall.

We conducted this audit in the Wage and Investment Division Headquarters Office in Atlanta, Georgia, and Wage and Investment Division offices in Philadelphia, Pennsylvania; New Carrollton, Maryland; and New York, New York. Audit work was also conducted in the Office of the Director, Research, Analysis and Statistics of Income in Washington, D.C.

The audit was conducted between April and August 2001 and in accordance with *Government Auditing Standards*. Detailed information on our audit objective, scope, and methodology is presented in Appendix I. Major contributors to the report are listed in Appendix II.

Weaknesses in the Sampling Strategy May Bias the Accuracy of the Results

The IRS made a good attempt to obtain a representative measure of its tax law assistance within the limitations of its current sampling strategy. The IRS implemented a sampling strategy that required quality reviewers to select, monitor, and assess the quality of responses to calls in “real” time during live calls. During the 2001 Filing Season, it reviewed 11,481 calls. The results provided the IRS with useful indicators of the quality of the responses it provided. However, plan design and implementation weaknesses in the current sampling strategy had the potential to bias the accuracy of the results. Also, managerial reviews were not conducted as required at critical periods during the 2001 Filing Season and were not performed in an independent setting.

According to a Professor of Decision Sciences¹⁰ that reviewed the sampling plan for the 2001 Filing Season, “At best, the sampling plan attempts to obtain ‘representative coverage’ of the quality of responses to tax law calls across the 15 tax sites every 2 weeks.” The sampling plan design attempted to ensure a certain number of calls were reviewed from all call sites at various times of the day. However, the plan did not include all hours or days of operation, used a sampling formula that likely resulted in a wider precision

¹⁰ Dr. Joseph Katz, PhD. Professor of Decision Sciences. Georgia State University.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

than reported, was designed in a mutually exclusive manner (i.e., if one event occurs, the other(s) cannot), and did not include r-mail calls. These weaknesses could bias the overall accuracy rate and precision for the sample, affecting the sample's reliability. We were unable to estimate the impact of the weaknesses. (The IRS' true accuracy for toll-free tax law assistance could be higher or lower than estimated or at 73.78 percent as reported.) To do this would require the re-creation of the sample, which is not possible.

Weaknesses existed in the design of the sampling plan

Our review identified several weaknesses in the design of the sampling plan. Specifically, the plan did not cover all hours and days of operation, used the simple random sampling formula for cluster sampling, and was designed in a "mutually exclusive" manner. Also, calls from taxpayers seeking tax law assistance that could not be answered during that initial call (r-mail) were not included. These design elements kept the IRS' sample from being truly random.

The plan did not cover all hours and days of operation. The IRS' toll-free tax law service was available 24 hours a day, 7 days a week during the 2001 Filing Season. However, the IRS selected calls for review only between the hours of 7:00 a.m. and 11:00 p.m. Eastern Standard Time, Monday through Saturday. Therefore, no calls received after 10:00 p.m. Central Standard Time, 9:00 p.m. Mountain Standard Time, or 8:00 p.m. Pacific Standard Time were monitored. Also, calls received on Sundays were not monitored.

Even during the 7:00 a.m. to 11:00 p.m. monitored time frame, the sampling plan did not cover all hours of operations for each call site. For example, one site that was open for 13 hours during the 7:00 a.m. to 11:00 p.m. time frame was only scheduled to be monitored up to 8 of the 13 hours that it was open.

The IRS estimated that 93 percent of the tax law calls came in during the hours reviewers were monitoring calls. Although the volume of calls outside of the time periods monitored was low in relation to the total received, the

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

ultimate effect on the sampling plan was that these calls did not have a chance of being selected.

The sampling plan did not include call site assignments for Saturdays. The sampling plan assigned reviewers to specific call sites. Assignments were made at varying times of the day on Monday through Friday, according to hours of operation of the calls sites and tours of duty of the reviewers. The sampling plan did not include Saturdays because sites that were available to answer tax law calls on Saturdays varied from week to week. During the week, reviewers were notified of the call sites that were operational each Saturday, but there was no systematic schedule for monitoring these call sites on Saturdays.

Using the simple random sampling formula for cluster sampling likely resulted in a wider precision. Under the cluster sampling method, the IRS grouped tax law calls within each call site into 1-hour sampling units or clusters and then selected calls within the hour to monitor for quality. The IRS used cluster sampling to select the calls to monitor; however, it used simple random sampling to determine the sample size (ideal number of calls required to be monitored) and to estimate the results. The sample size was determined based on the statistical simple random sampling formula.¹¹ The effect of applying the simple random sample formula to cluster sampling likely resulted in a wider precision.

The plan was designed in a mutually exclusive manner. Only one reviewer was scheduled to monitor a particular call site at a specific time. If the sampling plan was followed, other incoming calls that occurred at that call site during the same time period a reviewer was already monitoring a call could not be selected for review. As a result, all calls coming into that call site did not have an equal chance of selection for the sample.

¹¹ The established simple random sampling formula is $n = z^2 [p (1-p)/d^2]$. In this formula, p =historical accuracy, d =desired precision margin, and z =constant for confidence level. The IRS' sample size calculation was based on a historical accuracy rate of 71 percent, a confidence level of 90 percent, and a precision margin of 5 percent.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

Calls from taxpayers that were forwarded to the r-mail system were not included in the sample. There were occasions when a call received by a CSR was not answered initially (for example, for complex tax law subjects such as Individual Retirement Accounts, Capital Gains, and Sale of Residence). For these calls (r-mail), the CSR was to ask the taxpayer to leave a name and telephone number or electronic mail address so that the taxpayer could be provided a response within a few days. During the 2001 Filing Season, there were approximately 1 million taxpayer calls forwarded to r-mail.

As noted in a recent Treasury Inspector General for Tax Administration report,¹² even though these taxpayers sought assistance through the IRS' toll-free tax law telephone system, the quality assessment when the IRS later responded to these taxpayers was separately computed and did not factor into the overall estimate of accuracy for toll-free tax law telephone quality. Because the response is not immediate, the IRS does not believe this should be part of the toll-free tax law assistance measure for 'live' calls. After discussions with IRS management during this review, they changed the definition of the measure to include only 'live' toll-free tax law assistance.

Weaknesses existed in the implementation of the sampling plan

Once a sampling plan has been designed, it is critical that it be followed to ensure that it achieves the expected outcome. However, reviewers had latitude in selecting call sites and applications to monitor, had the ability to disconnect from a call at any time during monitoring with no systematic record that the call was selected, and did not monitor the required numbers of calls at critical times during the 2001 Filing Season.

Reviewers had latitude in selecting the site and application to monitor. Although the plan directed reviewers to a specific call site at a specific time, they had latitude in

¹² *Opportunities Exist to Improve the Performance Indicators Used to Convey Toll-Free Telephone Accuracy Accomplishments* (Reference Number: 2001-40-130, dated August 2001).

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

selecting specific calls to be monitored at the assigned site. For example, even though a reviewer was assigned two to four applications to monitor, the reviewer could impose his or her own preference to review one particular application over another. Also, if a call was not available to monitor at a site for his or her assigned applications, he or she was allowed to deviate to another call site.

By deviating from the sampling plan, reviewers could influence which calls were included in the sample and distort the true randomness of the sample. This conclusion is shared by the Professor of Decision Sciences¹³ who wrote, “The ‘experimental design’ or ‘controlled experiment’ approach adopted by the IRS attempts to limit the discretion of the reviewers, who are charged with the decision regarding which incoming call to monitor. Given a choice of calls to monitor, each reviewer can infuse his/her own preference bias into which call to select. This discretion can bias the overall quality estimates and precision estimates for the sample.”

The sampling plan was designed for reviewers to monitor calls at specific call sites at specific hours of the day. According to CQRS management, deviations from the assigned applications were allowed without prior managerial approval. However, deviations from the assigned call site were only to be to approved sites¹⁴ and then documented by the reviewers. Documentation requirements included annotating the time, the call site unable to be monitored, the substituted call site, and any applicable remarks.

We selected a statistically valid sample¹⁵ of 75 tax law calls monitored by reviewers during the 2001 Filing Season and

¹³ Dr. Joseph Katz, PhD.

¹⁴ CQRS analysts forwarded a weekly electronic mail message to reviewers indicating call sites with a low volume of calls monitored for use as approved call sites when call site deviation was necessary. Call site deviation was necessary at times when call sites were not available to answer calls, such as for severe weather conditions or short notice of low CSR staffing.

¹⁵ Sample was based on a 95 percent confidence level, a 5 percent precision margin, and a 5 percent expected error rate.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

reviewed the IRS' documentation¹⁶ pertaining to the call. Our sample showed that reviewers deviated from the assigned site 12 percent (9 times) of the time. When we apply this 12 percent to the total calls monitored by the CQRS, we estimate that 1,377 call site deviations may have occurred during the 2001 Filing Season. Reviewers did not document the reason for the deviations in six of the instances in our sample. The degree of latitude, even to approved sites, could affect the randomness of the sample.

Reviewers had the ability to disconnect from a call at anytime. Reviewers could disconnect from calls at any time for various reasons. For example, the reviewers were instructed to disconnect from a call if it was transferred outside the tax law area. However, no controls were in place to prevent reviewers from disconnecting from a call for reasons such as complexity of the issue, topic preference, or length of the call.

Each day, reviewers were to manually document the number of disconnected calls on information sheets maintained with their case review forms. However, CQRS management did not monitor the number of disconnected calls. We could not assess the frequency with which reviewers disconnected from calls because there was no way to systemically track the reviewer activity within the toll-free telephone system. This degree of latitude could also affect the randomness of the sample.

Reviewers did not monitor the required number of calls at critical times during the 2001 Filing Season. Once the ideal sample size was determined, the IRS evaluated its ability to meet the sampling plan based on resource assumptions. For example, one assumption was that the reviewers would be available an average of only 6.5 hours out of an 8.5 hour work day to monitor calls.¹⁷

¹⁶ IRS documentation consisted of the sampling plan, quality review database information, and associated daily information sheets.

¹⁷ The 6.5 hours takes training time, staff meeting time, lunch, and breaks into consideration. The 8.5-hour day includes a half-hour lunch break.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

For the 2001 Filing Season, the IRS calculated the ideal sample size¹⁸ at 223 calls per site and 3,345 calls nationwide (223 calls per site times 15 sites). The IRS resource assumptions for the 2001 Filing Season indicated that sufficient staffing was available to meet the sample size of 223 calls per call site per month for all 15 call sites (i.e., 3,345 calls nationwide per month).

Although the design of the sampling plan necessitated that at least 223 calls per month be monitored at each call site, the IRS did not meet the sampling plan at the call site level in three of four months of the 2001 Filing Season.¹⁹

Appendix IV (Figure 5) provides information for the planned and actual number of monitored calls per call site per month during the 2001 Filing Season.

While we could not substantiate the primary cause for not meeting the sampling plan, the IRS cited several reasons, such as low call volumes at certain times, problems with the software used to locate call traffic, and unavailability of reviewers during parts of the 2001 Filing Season.

The results by call site and on a nationwide basis provided the IRS with the opportunity to gauge the overall performance of each call site and implement changes targeted at improving performance, such as identifying training needs in specific areas. The assessment at the call site level was particularly important because not all call sites performed at the same level during the 2001 Filing Season. For example, during the month when the most calls were handled, one call site had a tax law accuracy rate of 80.28 percent (+/-4.48 percent), while another call site had a 60.27 percent tax law accuracy rate (+/-5.38 percent).

Although there was not a material effect on the nationwide accuracy rate when the monthly sample size was not met, meeting the sample size at the call site level was important because of the degree of risk associated with only sampling

¹⁸ Using a precision margin of +/-5 percent at a 90 percent confidence level and a 71 percent expected accuracy rate.

¹⁹ The required number of calls per site was not monitored for any of the 15 call sites in January, 11 of the 15 call sites in February, and 3 of the 15 call sites in April.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

parts of the entire population. When sample sizes fall below the required number, the degree of sampling error (precision) of the results will be wider or broader.

As stated in the background section of this report, IRS management must decide the maximum amount of error due to sampling that they will tolerate. IRS management selected a 5 percent sampling error. A wide precision in individual call site results may not provide meaningful information for them to make sound decisions about call site improvements. For example, during January 2001, one site had an accuracy rate of approximately 65.22 percent, +/-7.31 percent. This means that the true accuracy rate would lie somewhere between 57.91 percent and 72.53 percent – information that may not be meaningful to management. Appendix IV (Figure 6) provides call site accuracy and precision rates for January 2001.

Managerial reviews were not conducted at critical periods during the 2001 Filing Season

Managers in the CQRS had minimum monthly review requirements. We determined that these reviews of the work performed by the quality reviewers were not conducted as required in January, February, and March 2001. The manager of the quality reviewers was temporarily reassigned and a replacement was not designated.

Even though managerial reviews were not conducted as required, our analysis of IRS documentation of calls did not identify any material problems.

- In our review of the IRS' documentation for the 75 monitored calls previously mentioned, we determined that the reviewers' assessments of the accuracy of the responses to taxpayers were correct for 69 of 71 calls. For the remaining four calls, we could not determine the accuracy based on the limited documentation available on the call. However, it should be noted that our analysis of the 75 calls was dependent on the transcribed notes of the reviewers, as we could not re-create the actual taxpayer call.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

- In our review of a statistically valid sample of 68 of 1,020 edited case review forms on the tax law calls monitored,²⁰ we concluded that the reasons were appropriate for the changes.

One aspect of the managerial reviews consisted of the joint monitoring of live calls conducted side-by-side with the quality reviewers. This setting, not being independent, may have minimized the effectiveness of these managerial reviews in reducing some of the weaknesses identified in the current sampling strategy, such as deviating from the sampling plan and disconnecting from calls. Improving the managerial review process to conduct all required reviews in an independent setting could help to ensure the sampling plan is implemented as designed.

Errors in data on call volumes did not have a material impact

To ensure each call site was given the proper amount of weight in the overall results, the respective call volumes were used to “level the playing field.” These call volumes originated from an automated telephone routing system, and then were manually calculated and transcribed into the quality review database. Our review of the manual process that summarizes tax law call volumes at the call site level identified errors in the compilation of the data, but the errors were not material enough to affect the overall accuracy rate. The errors ranged from minimal in some call sites to overstating more than 25,000 calls in 2 call sites and understating more than 47,000 calls in one other. Although these errors did not change the overall accuracy rate when taken into account, there is the risk that errors, depending on their size, could have an effect. Since our discussion of these errors with IRS management, they have automated this process. We did not test the automated process to determine its reliability.

²⁰ Our sample of computerized records of case review forms was based on a 95 percent confidence level, a 5 percent precision margin, and a 5 percent expected error rate.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

In conclusion, statistical sampling provides a means to make conclusions about a population when only a sample of that population is reviewed. The IRS made a good attempt within the limitations of its current sampling strategy to obtain a representative measure of the quality of toll-free tax law assistance. Weaknesses resulting from the current sampling strategy, in terms of design and implementation, may bias the sample results. However, the exact impact cannot be quantified without re-creating the sample.

Recommendations

To improve the IRS' measure of accuracy under the current sampling strategy, the Commissioner, Wage and Investment Division, should:

1. Design the sampling plan to include all tax law calls in the population and randomly select from all hours of all call site operations.

Management's Response: IRS management stated, "The sampling plan we developed for the 2002 Filing Season covers all of the days and hours CSRs will provide tax law service. We have hired additional CQRS staff and authorized overtime to assure we execute the sample plan as designed. The sampling plan we developed offers the highest level of coverage by using the most random sampling our telephone system allows."

2. Ensure the sampling plan is implemented as designed, the latitude of reviewers to select and disconnect calls (in terms of both application and call site) is limited, and the required number of calls is reviewed at critical times during the filing season.

Management's Response: IRS management stated, "The 2002 version of the plan does not allow reviewers to select call sites and applications. Instead, reviewers are assigned to specific sites and applications. Now we check daily information sheets prepared by the reviewers to ensure they follow the sample plan. If we must deviate from the plan because of site operating conditions or closures, we follow a predetermined process to maintain random selection. In addition, we

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

will select a separate sample of calls and compare them to the sample plan to ensure reviewers are not deviating from their hourly assignments. No telephone systems exist that can prevent the user from disconnecting from a call. The telephone equipment the CQRS staff uses cannot track disconnected calls.

We are using a status report to compare the projected sample per site and the actual samples taken. If the projection is not met, we will determine the reason(s) and take corrective action. We have scheduled staff resources, including overtime, to ensure we consistently meet sample sizes.”

3. Conduct the required managerial reviews of the work performed by quality reviewers and perform the on-line, joint-monitoring aspect of these reviews remotely so that it can be done in an independent setting. These reviews should help ensure that the sampling plan is executed as designed.

Management’s Response: IRS management stated, “We cannot implement this recommendation due to telephone system limitations. The telephone system used by CQRS does not allow simultaneous remote call monitoring. CQRS managers conduct post reviews of their reviewers’ work by analyzing their data collection instruments and call notes. This review supplements the side-by-side joint monitoring that is conducted by each manager. We have strengthened controls to ensure that each CQRS manager meets minimum monitoring standards.”

To ultimately address the design limitations of the current sampling strategy the Commissioner, Wage and Investment Division, should:

4. Develop a system to measure the accuracy of tax law assistance that ensures that all tax law calls are included in the population, the selection of calls is truly random, and the sampling plan is implemented as designed. One way to accomplish this would be to institute an automated call recording system that would also provide a true random method of

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

selecting calls from the entire population of calls to the IRS' toll-free tax law assistance telephone system.

Management's Response: IRS management stated, "While we agree that an automated call recording system is the best way to achieve random call selection, we are not technologically able to do so now. We are aggressively pursuing automated call recording. Meantime, we continue to improve our current processes to achieve as high a level of statistical reliability as possible."

Detailed Objective, Scope, and Methodology

The overall objective of this review was to determine if the Internal Revenue Service (IRS) reliably measured the accuracy of responses that millions of taxpayers experienced when they called the IRS for tax law assistance.

We reviewed the process the IRS uses to measure the accuracy of its tax law assistance and determined the statistical reliability of each component of the process. This included reviewing actual case files from quality reviews at the Centralized Quality Review Site (CQRS), reviewing the reliability of the sampling methodology used to measure the accuracy, and reviewing the reliability of the individual components that comprised the measure of accuracy.

To accomplish our objective, we:

- I. Reviewed the current sampling plan to determine, if executed properly, whether it would result in a statistically valid estimate. (Design of the Sample)
 - A. Interviewed CQRS and Statistics of Income (SOI) staff to determine the purpose of the sample, how it was developed, and the attributes of the population used.
 - B. Interviewed SOI staff and reviewed industry practices to determine the basis for the confidence and precision levels set by the IRS.
 - C. Researched industry practices and consulted with contracted statistician to determine whether the appropriate and necessary elements or attributes of the population were addressed in the sampling plan.
 - D. Analyzed quality review database information for the 2000 and 2001 Filing Seasons and compared the results to the 2001 Filing Season sampling plan for indications of bias.
- II. Evaluated whether the sampling plan was executed by the CQRS as designed by the SOI.
 - A. Reviewed quality review database and call volume data to determine if the CQRS met the sampling plan.
 - B. Interviewed CQRS staff to determine reasons why the sampling plan was not met.
 - C. Interviewed CQRS staff to determine reasons for deviations from the sampling plan. Also, reviewed a statistically valid random sample of review documentation for the 75 calls from II.G. to determine if the reviewers had deviated from the sampling plan. The sample was selected from a universe of 11,459 calls monitored during the 2001 Filing Season, at a confidence level of 95 percent, an error rate of 5 percent, and a precision margin of 5 percent.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

- D. Analyzed quality review database information to determine if reviewers were monitoring sites and applications outside of the sampling plan.
 - E. Interviewed CQRS staff and reviewed training material to determine if reviewers had required skill sets.
 - F. Interviewed CQRS management and reviewed performance review documentation to determine if there was proper oversight.
 - G. Reviewed a statistically valid random sample of review documentation for 75 calls to determine the overall accuracy of the determination, if calls could be reconstructed, and if calls were accurately transcribed. The sample was selected from a universe of 11,459 calls monitored during the 2001 Filing Season, at a confidence level of 95 percent, an error rate of 5 percent, and a precision margin of 5 percent.
 - H. Reviewed a statistically valid sample of 68 edited cases for review to determine reasons for the edits. The sample was selected from a universe of 1,020 edited records during the 2001 Filing Season at a confidence level of 95 percent, an error rate of 5 percent, and a precision margin of 5 percent.
- III. Evaluated whether results of the sampling plan were compiled correctly.
- A. Interviewed operations program analysts at the IRS National Headquarters and SOI management staff to determine how the sampling plan results were compiled and how the mathematical statistical formulas were used in the compilation.
 - B. Determined if the information used in the calculation of the quality rates and precision margins by the quality database was accurate by using original source data to test the individual components of the quality calculations.
 - C. Verified that the calculations used by the quality database to report the individual call site and national rollup quality rates and precision margins were accurate, based on actual data previously entered into the database, by recalculating the individual and national results.
 - D. Determined by discussion with SOI management staff, IRS operations program analysts, and CQRS management staff what validation the IRS performed on the quality rate reporting methodology in order to ensure the accuracy of the data and the statistical validity of the results calculated by the quality database.
 - E. Determined how certain attributes of the toll-free tax law call population, such as calls received/monitored by call site and calls received/monitored by application, were represented in the overall final sample selected for purposes of reporting the IRS national toll-free tax law quality rate during the 2001 Filing Season.

Major Contributors to This Report

Michael R. Phillips, Acting Assistant Inspector General for Audit (Wage and Investment
Income Programs)
Susan Boehmer, Director
Stan Rinehart, Director
Patricia Lee, Audit Manager
Anthony Anneski, Senior Auditor
Deborah Carter, Senior Auditor
Gregory Dix, Senior Auditor
Kathleen Hughes, Senior Auditor
Doris Hynes, Senior Auditor
Sharla Robinson, Senior Auditor
Jerry Douglas, Auditor
Andrea McDuffie, Auditor
Geraldine Vaughn, Auditor

**The Internal Revenue Service Could Improve Its Process to More
Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance**

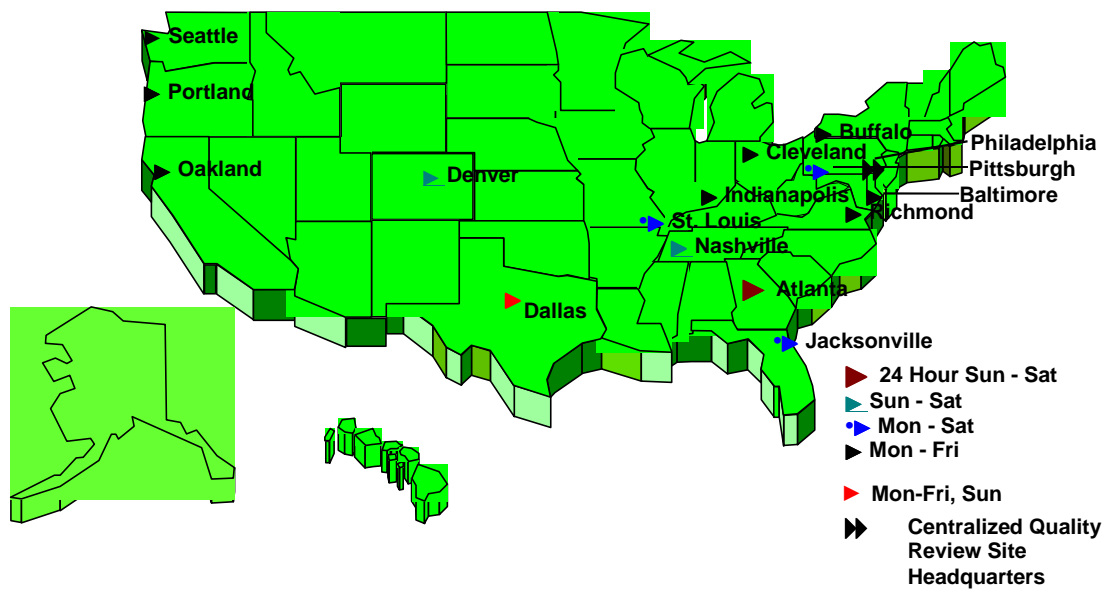
Appendix III

Report Distribution List

Commissioner N:C
Commissioner, Small Business/Self-Employed Division S
Director, Customer Account Services W:CAS
Director, Strategy and Finance W:S
Director, Research, Analysis, and Statistics of Income N:ADC:R
Chief, Customer Liaison S:COM
Chief Counsel CC
National Taxpayer Advocate TA
Director, Legislative Affairs CL:LA
Director, Office of Program Evaluation and Risk Analysis N:ADC:R:O
Office of Management Controls N:CFO:F:M
Audit Liaisons:
 Commissioner, Wage and Investment Division W
 Director, Customer Account Services W:CAS
 Director, Research, Analysis, and Statistics of Income N:ADC:R

Tables and Charts

Figure 1
Internal Revenue Service Tax Law Call Site Locations



Source: Internal Revenue Service Call Sites

**The Internal Revenue Service Could Improve Its Process to More
Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance**

Figure 2

Hours of Operation for the 15 Internal Revenue Service Tax Law Call Sites

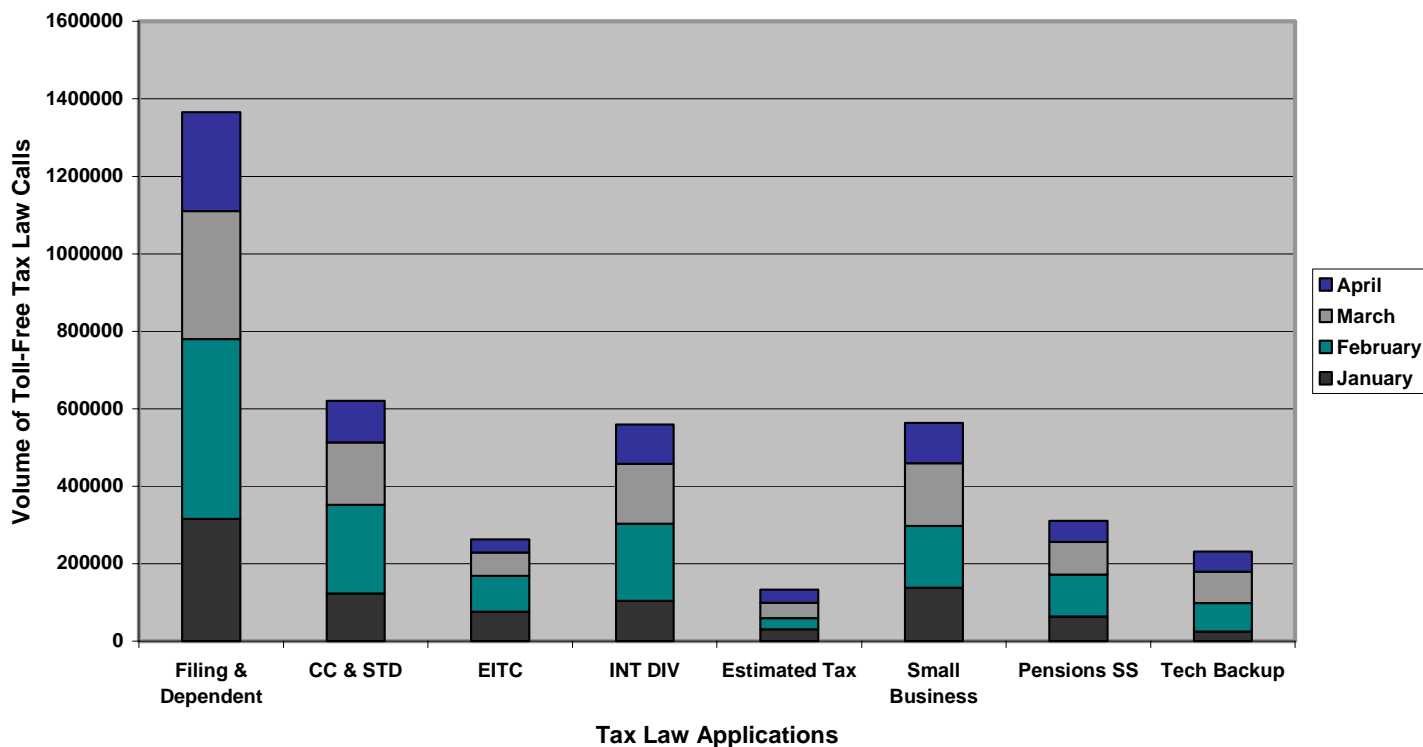
Sunday – Saturday	Atlanta	24 Hours	
	Denver	9:00A – 7:00P	
	Nashville	7:00A – 8:00P	Mon - Fri
		7:00A – 3:00P	Sat, Sun
Monday – Saturday	Jacksonville	6:30A – 8:30P	Mon - Fri
		10:30A – 8:30P	Sat
	Pittsburgh	6:00A – 4:30P	Mon - Fri
		8:00A – 4:30P	Sat
	St. Louis	7:30A – 5:30P	Mon - Fri
		6:00A – 3:00P	Sat
Monday - Friday	Baltimore	6:30A – 4:30P	
	Buffalo	6:30A – 10:00P	
	Cleveland	7:00A – 4:30P	
	Dallas	7:00A – 11:30P	Sun 11:30A-11:30P
	Indianapolis	8:00A – 8:00P	
	Oakland	10:00A – 10:00P	
	Portland	9:30A – 10:00P	
	Richmond	6:30A – 7:00P	
	Seattle	9:30A – 10:00P	

Source: Internal Revenue Service Call Sites

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

Figure 3

**Internal Revenue Service Toll-Free Tax Law Call Volumes
by Application (2001 Filing Season)**



Source: Internal Revenue Service (IRS) Automated Telephone Routing System

Abbreviations

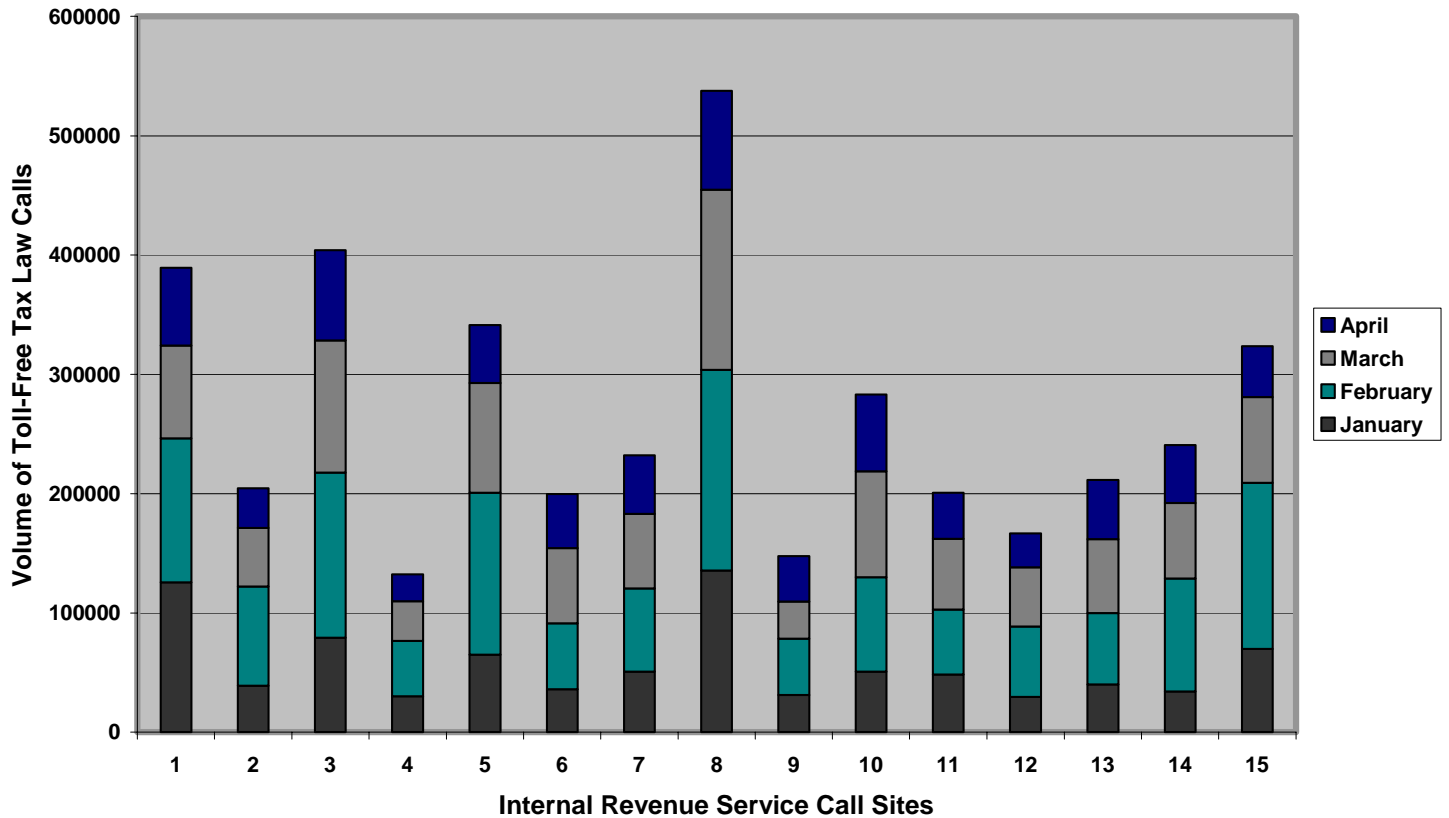
CC = Charitable Contributions
 STD = Standard Deduction
 EITC = Earned Income Tax Credit
 INT DIV = Interest and Dividend Income
 SS = Social Security
 Tech Backup = Technical Backup

The volume of toll-free tax law calls handled by tax law application by the IRS varied. The volume of calls handled for the Filing and Dependent tax law application was the highest for each month from January through April.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

Figure 4

**Internal Revenue Service Toll-Free Tax Law Call Volumes
by Call Site (2001 Filing Season)**



Source: Internal Revenue Service (IRS) Automated Telephone Routing System

The volume of toll-free tax law calls handled by each IRS call site varied.

**The Internal Revenue Service Could Improve Its Process to More
Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance**

Figure 5

**Internal Revenue Service Tax Law Calls Monitored
Compared to Planned¹ for Call Sites (2001 Filing Season)**

Call Sites	January		February		March		April	
	Calls Monitored	Over (Under) Planned	Calls Monitored	Over (Under) Planned	Calls Monitored	Over (Under) Planned	Calls Monitored	Over (Under) Planned
1	165	(58)	224	1	313	90	162	50
2	162	(61)	211	(12)	268	45	109	(3)
3	150	(73)	199	(24)	280	57	137	25
4	154	(69)	200	(23)	273	50	106	(6)
5	140	(83)	169	(54)	263	40	151	39
6	154	(69)	166	(57)	242	19	127	15
7	151	(72)	231	8	274	51	134	22
8	179	(44)	213	(10)	305	82	140	28
9	151	(72)	214	(9)	255	32	131	19
10	115	(108)	185	(38)	265	42	114	2
11	199	(24)	232	9	332	109	156	44
12	136	(87)	147	(76)	235	12	110	(2)
13	145	(78)	193	(30)	277	54	153	41
14	103	(120)	185	(38)	262	39	132	20
15	182	(41)	262	39	339	116	119	7
Totals	2286	(1059)	3031	(314)	4183	838	1981	301

Source: Internal Revenue Service (IRS) Quality Review Database

In January, February, and April, the IRS did not meet the sampling plan for all of its call sites.

¹ The IRS planned to monitor 223 calls per site, per month for January, February, and March (223 calls X 15 sites = 3345). The IRS planned to monitor only 112 calls per site through mid-April because the Centralized Quality Review Site (CQRS) discontinued its monitoring operation for the remainder of April to complete the end of the filing season processing.

**The Internal Revenue Service Could Improve Its Process to More
Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance**

Figure 6
Accuracy Rates and Precision Margins for Internal Revenue Service Call Sites
(January 2001)

Call Site	Accuracy Rate	Precision Margin
1	62.42%	+/- 6.20%
2	72.22%	+/- 5.79%
3	64.00%	+/- 6.45%
4	68.83%	+/- 6.14%
5	62.14%	+/- 6.74%
6	71.43%	+/- 5.99%
7	75.50%	+/- 5.76%
8	81.01%	+/- 4.82%
9	72.19%	+/- 6.00%
10	65.22%	+/- 7.31%
11	73.87%	+/- 5.12%
12	74.26%	+/- 6.17%
13	72.41%	+/- 6.11%
14	71.84%	+/- 7.29%
15	64.29%	+/- 5.84%

Source: Internal Revenue Service (IRS) Quality Review Database

During January 2001 when the sampling plan was not met for the IRS call sites (refer to Figure 5), the precision margin (sampling error) for many of the call sites exceeded the 5 percent sampling error desired by the IRS (based on the 5 percent sampling error the IRS used in its simple random sample formula to determine sample size).

Common Sampling Techniques

- **Random sampling** relies entirely on chance. By this method, *every* item in the population is given an *equal* chance of being included in the sample. For example, each call answered by the Internal Revenue Service (IRS) must have an equal chance of selection into its sample or the method is not truly random. Unless each item is selected by a purely random technique, there is no way of measuring later how accurately the sample reflects the characteristics of the population from which it was selected. One major benefit of statistical random sampling is that it permits objective measurement of the reliability of the results. Established mathematical formulas are used to determine the proper sample size necessary to measure the reliability of the results. This sampling method is used when conclusions about the entire population are to be presented. Random sampling methods include simple random, stratified and cluster sampling.¹
- **Judgmental sampling** (also known as non-random) does not rely on the principal of chance to select the sample. In this type of sample, the sampler's best judgment (possibly based on past experience) is used in selecting those items for the sample that are believed to give a representative picture of the universe. Although a judgmental sample may give a good indication of the population, this type of sample does not lend itself to analysis by standard statistical methods such as assessing the reliability of the results. Judgmental samples are also difficult to defend against challenges regarding their validity and reliability. Therefore, the results cannot be used to present conclusions about the whole.²
- **Convenience sampling** (also known as spot-check sampling) is neither a judgmental nor a statistical random (probability) sample. It differs from statistical random sampling in that the items usually included in the sample are "grab" items. This type of sample rests on the illusion that no rule is the best rule for obtaining a representative sample. There is neither a control to assure a known chance of selection nor a system of considered judgment.³

¹ Mandel & Laessig, p. 174, 184-205, 233, 259.

² Mandel & Laessig, p. 202, 205, 479.

³ Mandel & Laessig p. 203.

The Internal Revenue Service Could Improve Its Process to More
Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

Appendix VI

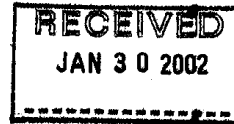
Management's Response to the Draft Report




DEPUTY COMMISSIONER

DEPARTMENT OF THE TREASURY
INTERNAL REVENUE SERVICE
WASHINGTON, D.C. 20224

January 29, 2002



MEMORANDUM FOR TREASURY INSPECTOR GENERAL FOR TAX
ADMINISTRATION

FROM: Bob Wenzel 
Deputy Commissioner of Internal Revenue

SUBJECT: Draft Audit Report – "The Internal Revenue Service Could
Improve Its Process to More Reliably Measure the Accuracy of
Its Toll-Free Tax Law Assistance" (Audit # 200140039)

Measuring the accuracy of tax law information we provide to toll-free callers is key to continuous improvement. The Centralized Quality Review Site (CQRS) replaced the test question technique with on-line, live monitoring of real taxpayer questions and assistor responses.

The process is supported by a user-friendly Quality Review Database (QRDB) that is accessible to employees involved with measuring and improving our services. As soon as the CQRS staff enters feedback into the QRDB, it is available to correct problems on the spot. We have used the information from more than 11,000 monitored calls during the 2001 Filing Season to develop short and long-term improvement actions.

I believe the statistical data is sound and reflects our performance. It is influenced by the systemic constraints shown in your report. The statistical experts acknowledged during the audit that our efforts to compensate for these limitations were positive and reasonable. We will use your recommendations to further improve the statistical validity of the sampling plan and its execution. I agree, the best solution would be an automated recording process, which we are actively pursuing.

Our comments on your recommendations follow:

RECOMMENDATION 1

To improve the IRS' measure of accuracy under the current sampling strategy, the Commissioner, Wage and Investment Division, should:

Design the sampling plan to include all tax law calls in the population and randomly select from all hours of call site operations.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

2

ASSESSMENT OF CAUSE

The report states the sampling plan did not cover all hours and days of operation, used the simple random sampling formula for a cluster sampling, and was designed in a "mutually exclusive" manner.

CORRECTIVE ACTION

The sampling plan we developed for the 2002 Filing Season covers all of the days and hours Customer Service Representatives (CSR) will provide tax law service. We have hired additional CQRS staff and authorized overtime to assure we execute the sample plan as designed. The sampling plan we developed offers the highest level of coverage by using the most random sampling our telephone system allows.

IMPLEMENTATION DATE:

Completed

RESPONSIBLE OFFICIALS

Director, Customer Account Services
Commissioner, Wage and Investment Division

RECOMMENDATION 2

To improve the IRS' measure of accuracy under the current sampling strategy, the Commissioner, Wage and Investment Division, should:

Ensure the sampling plan is implemented as designed, the latitude of reviewers to select and disconnect calls (in terms of both application and call site) is limited, and the required number of calls is reviewed at critical times during the filing season.

ASSESSMENT OF CAUSE

We did not follow the sampling plan closely enough for complete statistical assurance. Reviewers had limited latitude to select call sites and applications to monitor and to disconnect from a call during monitoring with no systemic record that the call was selected. Reviewers also did not monitor the required number of calls at critical times during the 2001 Filing Season.

CORRECTIVE ACTION

The 2002 version of the plan does not allow reviewers to select call sites and applications. Instead, reviewers are assigned to specific sites and applications. Now we check daily information sheets prepared by the reviewers to ensure they follow the sample plan. If we must deviate from the plan because of site operating conditions or closures, we follow a predetermined process to maintain random selection. In addition, we will select a separate sample of calls and compare them to the sample plan to ensure reviewers are not deviating from their hourly assignments. No telephone systems exist that can prevent the user from disconnecting from a call. The telephone equipment the CQRS staff uses cannot track disconnected calls.

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

3

We are using a status report to compare the projected sample per site and the actual samples taken. If the projection is not met, we will determine the reason(s) and take corrective action. We have scheduled staff resources, including overtime, to ensure we consistently meet sample sizes.

IMPLEMENTATION DATE

Completed

RESPONSIBLE OFFICIALS

Director, Customer Account Services
Commissioner, Wage and Investment Division

RECOMMENDATION 3

To improve the IRS' measure of accuracy under the current sampling strategy, the Commissioner, Wage and Investment Division, should:

Conduct the required managerial reviews of the work performed by the quality reviewers and perform the on-line, joint monitoring aspect of these reviews remotely, so it can be done in an independent setting. These reviews should help ensure that the sampling plan is executed as designed.

ASSESSMENT OF CAUSE

The report states CQRS managers had minimum monthly review requirements, which they did not meet in January, February, and March 2001. When the reviewers' manager was temporarily reassigned, a replacement was not designated. TIGTA did not identify any material problems based on their analysis of call documentation during this period. The report notes managerial monitoring of live calls is conducted side-by-side with the reviewers. This setting is not independent, and it may have minimized the effectiveness of these managerial reviews in reducing some of the weaknesses identified in the current sampling strategy, such as deviating from the sampling plan and disconnecting from calls.

CORRECTIVE ACTION

We cannot implement this recommendation due to telephone system limitations. The telephone system used by CQRS does not allow simultaneous remote call monitoring. CQRS managers conduct post reviews of their reviewers' work by analyzing their data collection instruments and call notes. This review supplements the side-by-side joint monitoring that is conducted by each manager. We have strengthened controls to ensure that each CQRS manager meets minimum monitoring standards.

IMPLEMENTATION DATE:

Not Applicable

RESPONSIBLE OFFICIALS

Director, Customer Account Services
Commissioner, Wage and Investment Division

The Internal Revenue Service Could Improve Its Process to More Reliably Measure the Accuracy of Its Toll-Free Tax Law Assistance

4

RECOMMENDATION 4

To ultimately address the design limitations of the current sampling strategy, the Commissioner, Wage and Investment Division, should:

Develop a system to measure the accuracy of tax law assistance that ensures that all tax law calls are included in the population, the selection of the calls is truly random, and the sample plan is implemented as designed. One way to accomplish this would be to institute an automated call recording system that would also provide a true random method of selecting calls from the entire population of calls to the IRS' toll-free tax law assistance telephone system.

ASSESSMENT OF CAUSE

The limitation of current telephone technology affects our sampling strategy. Because of telephone system limitations, we have to monitor calls live. IRS management and the CQRS staff have tried to minimize the impact of these limitations.

CORRECTIVE ACTION

While we agree that an automated call recording system is the best way to achieve random call selection, we are not technologically able to do so now. We are aggressively pursuing automated call recording. Meantime, we continue to improve our current processes to achieve as high a level of statistical reliability as possible.

IMPLEMENTATION DATE:

Not Applicable

RESPONSIBLE OFFICIALS

Director, Customer Account Services
Commissioner, Wage and Investment Division

If you have any questions, please call John Dalrymple, Commissioner, Wage and Investment Division, at (202) 622-6860 or Ron Watson, Director, Customer Account Services at (404) 338-8910.